

## EAST Search History

Ref #	Hits	Search Query	DBs	Default Operator	Plurals	Time Stamp
L1	17	(bi\$2arsen\$ or (bi! or bis! or di!) adj (arsen\$ or arsa)).clm.	US-PGPUB; USPAT	OR	ON	2006/06/29 10:13
L2	8	(tetra\$1cystein\$ or tetra adj cystein\$ or cxxcc).clm.	US-PGPUB; USPAT	OR	ON	2006/06/29 10:13

Checked L1, L2  
JRU  
6-29-2006

# Library Sequence Search History

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

Checked  
JSR  
6-29-2006

OM protein - protein search, using sw model

Run on: June 19, 2006, 17:24:38 ; Search time 203 Seconds  
(without alignments)  
77.464 Million cell updates/sec

Title: US-10-772-164-1  
Perfect score: 101  
Sequence: 1 WEAAAREACCRECCARA 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database : UniProt\_7.2:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	% Query		Match Length	DB ID	Description
	Score	Match			
1	53	52.5	157	1	VE6_HP12 P36803 human papil
2	51	50.5	243	2	Q46MX5_RALEJ Q46mx5 ralstonia e
3	51	50.5	278	2	Q6IGH0_DROME Q6igh0 drosophila
4	51	50.5	502	2	Q9BGM9_9MAMM Q9bgm9 tachyglossu
5	51	50.5	1408	2	Q381X2_9TRYP Q381x2 trypanosoma
6	50	49.5	589	2	Q2JB13_9ACTO Q2jbl3 frankia sp.
7	50	49.5	1370	1	ZMYM3_HUMAN Q14202 homo sapien
8	50	49.5	1370	1	ZMYM3_MOUSE Q9jlm4 mus musculu
9	49.5	49.0	602	2	Q75NZ5_CHLRE Q75nz5 chlamydomon
10	49	48.5	115	1	ALK1_PIG P22298 sus scrofa
11	49	48.5	155	2	Q9PXB1_HP108 Q9pxb1 human papil
12	49	48.5	168	1	VE6_HP21 P28832 human papil
13	49	48.5	1067	2	Q4QFE4_LEIMA Q4qfe4 leishmania
14	48.5	48.0	390	2	Q4S604_TETNG Q4s604 tetraodon n
15	48	47.5	62	2	Q4PN38_IXOSC Q4pn38 ixodes scap

16	48	47.5	131	1	ALK1_MOUSE	P97430	mus musculu
17	48	47.5	131	2	Q548X8_MOUSE	Q548x8	mus musculu
18	48	47.5	157	2	O40617_HPVR7	O40617	human papil
19	48	47.5	157	2	Q81986_HPVO5	Q81986	human papil
20	48	47.5	157	2	Q913V6_9PAPI	Q913v6	human papil
21	47.5	47.0	525	2	Q64FQ2_ARATH	Q64fq2	arabidopsis
22	47.5	47.0	676	2	O48785_ARATH	O48785	arabidopsis
23	47	46.5	88	2	Q62H93_BURMA	Q62h93	burkholderi
24	47	46.5	101	2	Q4IVT4_AZOVI	Q4ivt4	azotobacter
25	47	46.5	131	2	Q9R0Z8_RAT	Q9r0z8	rattus norv
26	47	46.5	156	1	VE6_HP47	P22422	human papil
27	47	46.5	171	1	VE6_HP14	P28830	human papil
28	47	46.5	181	2	Q8VMH1_PSEPU	Q8vmh1	pseudomonas
29	47	46.5	193	1	KR415_HUMAN	Q9byq5	homo sapien
30	47	46.5	210	1	KRA47_HUMAN	Q9byr0	homo sapien
31	47	46.5	219	2	Q52396_PSEST	Q52396	pseudomonas
32	47	46.5	330	2	Q3E2M8_CHLAU	Q3e2m8	chloroflexu
33	47	46.5	399	2	Q3INU5_NATPD	Q3inu5	natronomona
34	47	46.5	438	2	Q341E3_RHOPA	Q341e3	rhodopseudo
35	47	46.5	1175	2	Q4P5X7_USTMA	Q4p5x7	ustilago ma
36	46	45.5	80	1	IBB4_LONCA	P16343	lonchocarpu
37	46	45.5	88	2	Q52509_PSESX	Q52509	pseudomonas
38	46	45.5	95	2	Q4CKP2_TRYCR	Q4ckp2	trypanosoma
39	46	45.5	100	2	Q37B06_RHOPA	Q37b06	rhodopseudo
40	46	45.5	129	1	KRA56_HUMAN	Q6l8g9	homo sapien
41	46	45.5	161	2	Q8MZ55_DROME	Q8mz55	drosophila
42	46	45.5	166	1	VE6_HP19	P36806	human papil
43	46	45.5	186	1	KRA45_HUMAN	Q9byr2	homo sapien
44	46	45.5	191	2	Q28583_SHEEP	Q28583	ovis aries
45	46	45.5	203	2	Q3VGI5_9SPHN	Q3vgi5	sphingopyxi
46	46	45.5	232	2	Q2I0E2_ORYSA	Q2i0e2	oryza sativ
47	46	45.5	298	2	Q65T35_MANSM	Q65t35	mannheimia
48	46	45.5	412	2	P91666_DROME	P91666	drosophila
49	46	45.5	465	1	HYIN2_BRAJA	P19922	bradyrhizob
50	46	45.5	491	2	Q4T2B4_TETNG	Q4t2b4	tetraodon n

#### ALIGNMENTS

#### RESULT 1

VE6\_HP12

ID VE6\_HP12 STANDARD; PRT; 157 AA.  
AC P36803;  
DT 01-JUN-1994, integrated into UniProtKB/Swiss-Prot.  
DT 01-JUN-1994, sequence version 1.  
DT 07-FEB-2006, entry version 27.  
DE Protein E6.  
GN Name=E6;  
OS Human papillomavirus type 12.  
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;  
OC Betapapillomavirus.  
OX NCBI\_TaxID=10604;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].  
RX MEDLINE=94265501; PubMed=8205838;  
RA Delius H., Hofmann B.;  
RT "Primer-directed sequencing of human papillomavirus types.";

```

RL   Curr. Top. Microbiol. Immunol. 186:13-31(1994).
CC   -!- FUNCTION: Transcriptional transactivator. Binds double stranded
CC       DNA (By similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear matrix-associated (By similarity).
CC   -!- SIMILARITY: Belongs to the papillomaviruses E6 protein family.
CC   -----
CC   Copyrighted by the UniProt Consortium, see http://www.uniprot.org/terms
CC   Distributed under the Creative Commons Attribution-NoDerivs License
CC   -----
DR   EMBL; X74466; CAA52496.1; -; Genomic_DNA.
DR   PIR; S36538; S36538.
DR   InterPro; IPR001334; E6.
DR   Pfam; PF00518; E6; 1.
KW   Activator; DNA-binding; Early protein; Metal-binding; Nuclear protein;
KW   Transcription; Transcription regulation; Zinc; Zinc-finger.
FT   CHAIN           1       157       Protein E6.
FT                                     /FTId=PRO_0000133332.
FT   ZN_FING         39       75       Potential.
FT   ZN_FING        112      148       Potential.
SQ   SEQUENCE      157 AA;  17984 MW;  E9EC735537733FDC CRC64;

```

```

Query Match          52.5%;  Score 53;  DB 1;  Length 157;
Best Local Similarity 53.3%;  Pred. No. 14;
Matches      8;  Conservative    1;  Mismatches      6;  Indels      0;  Gaps      0;

```

```

Qy      1 WEAAAREACCRECCA 15
        |:      |||| |||
Db      63 WKGHFVTACCRSCCA 77

```

Search completed: June 19, 2006, 17:39:02  
Job time : 243 secs

OM protein - protein search, using sw model

Run on: June 19, 2006, 17:32:24 ; Search time 24 Seconds  
(without alignments)  
68.154 Million cell updates/sec

Title: US-10-772-164-1  
Perfect score: 101  
Sequence: 1 WEAAAREACCRECCARA 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database : PIR\_80:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	% Query		Length	DB	ID	Description
	Score	Match				
1	53	52.5	157	2	S36538	E6 protein - human
2	49	48.5	115	2	A36113	antileukoproteinas
3	47.5	47.0	676	2	G84663	hypothetical prote
4	47	46.5	156	1	W6WL47	E6 protein - human
5	46	45.5	166	2	S36485	E6 protein - human
6	46	45.5	191	2	I46412	keratin KAP5.4 - s
7	46	45.5	465	2	S05311	indoleacetamide hy
8	46	45.5	498	2	A48203	interleukin-14 pre
9	46	45.5	571	2	S69210	protein kinase cak
10	46	45.5	1430	2	T34516	hypothetical prote
11	45	44.6	61	2	E82580	hypothetical prote
12	45	44.6	155	1	W6WL8	E6 protein - human
13	45	44.6	157	1	W6WL5	E6 protein - human

14	45	44.6	157	1	W6WLB5	E6 protein - human
15	45	44.6	273	2	A43862	29K peripheral mem
16	45	44.6	369	2	G75460	hypothetical prote
17	44	43.6	161	2	S36491	E6 protein - human
18	44	43.6	186	2	A45910	ultra-high-sulfur
19	44	43.6	188	2	JC6547	high sulfur protei
20	44	43.6	204	2	T08072	proteinase inhibit
21	44	43.6	251	2	AH3413	nitrogen fixation
22	44	43.6	254	2	B84901	hypothetical prote
23	44	43.6	299	2	C97102	hypothetical prote
24	44	43.6	370	1	S57347	Ca2+/calmodulin-de
25	44	43.6	374	1	S50193	Ca2+/calmodulin-de
26	44	43.6	496	2	F75257	hypothetical prote
27	44	43.6	994	2	A48849	Ca2+-transporting
28	44	43.6	1001	1	PWRBFC	Ca2+-transporting
29	44	43.6	1121	2	S30862	DNA dependent ATPa
30	43.5	43.1	126	2	I46489	cysteine-rich hair
31	43	42.6	169	1	S18946	ultra high-sulfur
32	43	42.6	217	2	T33353	hypothetical prote
33	43	42.6	221	2	C34768	ORF2 protein - Orf
34	43	42.6	233	2	S67947	alkyl hydroperoxid
35	43	42.6	399	2	B24698	formate dehydrogen
36	43	42.6	689	2	T08988	cadmium-transporti
37	43	42.6	711	2	A85352	cadmium-transporti
38	43	42.6	976	2	D96714	DNA-directed RNA p
39	42.5	42.1	931	2	H96527	protein F27J15.16
40	42	41.6	122	2	JC6548	high sulfur protei
41	42	41.6	223	2	B38346	ultra-high-sulfur
42	42	41.6	230	2	A38346	ultra-high-sulfur
43	42	41.6	247	2	T17311	hypothetical prote
44	42	41.6	327	2	C86452	protein F6N18.11 [
45	42	41.6	1212	2	B82809	exodeoxyribonuclea
46	42	41.6	2037	2	T16881	hypothetical prote
47	41	40.6	67	2	T37199	hypothetical prote
48	41	40.6	151	2	S60314	hair keratin cyste
49	41	40.6	164	2	T24272	hypothetical prote
50	41	40.6	169	2	T06062	hypothetical prote

# ALIGNMENTS

## RESULT 1

S36538

E6 protein - human papillomavirus type 12

C;Species: human papillomavirus type 12

C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004

C;Accession: S36538

R;Delius, H.; Hofmann, B.

submitted to the EMBL Data Library, August 1993

A;Description: Primer-directed sequencing of human papillomavirus types.

A;Reference number: S36469

A;Accession: S36538

A;Molecule type: DNA

A;Residues: 1-157 <DEL>

A;Cross-references: UNIPROT:P36803; UNIPARC:UPI00001383B8; EMBL:X74466;

NID:g396910; PIDN:CAA52496.1; PID:g396911

C;Superfamily: papillomavirus E6 protein

C;Keywords: DNA binding; early protein; nucleus; zinc finger

Query Match 52.5%; Score 53; DB 2; Length 157;  
Best Local Similarity 53.3%; Pred. No. 2.9;  
Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 WEAAAAREACCRECCA 15  
|: ||| |||  
Db 63 WKGHFVTACCRSCCA 77

Search completed: June 19, 2006, 17:39:33  
Job time : 40 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2006, 17:56:29 ; Search time 13 Seconds  
(without alignments)  
29.497 Million cell updates/sec

Title: US-10-772-164-1  
Perfect score: 101  
Sequence: 1 WEAAAREACCRECCARA 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 96747 seqs, 22556637 residues

Total number of hits satisfying chosen parameters: 96747

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database : Published\_Applications\_AA\_New:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		Length	DB	ID	Description
		Match					
1	48	47.5		4440	6	US-10-196-749-525	Sequence 525, App
2	45.5	45.0		139	6	US-10-449-902-42349	Sequence 42349, A
3	45	44.6		1129	6	US-10-527-411-42	Sequence 42, Appl
4	45	44.6		1129	6	US-10-527-411-48	Sequence 48, Appl
5	45	44.6		1129	6	US-10-527-411-52	Sequence 52, Appl
6	45	44.6		1129	6	US-10-527-411-56	Sequence 56, Appl
7	45	44.6		1132	6	US-10-527-411-46	Sequence 46, Appl
8	45	44.6		1894	6	US-10-196-749-97	Sequence 97, Appl
9	44	43.6		105	6	US-10-953-349-39499	Sequence 39499, A



10	44	43.6	449	6	US-10-953-349-8402	Sequence 8402, Ap
11	44	43.6	449	6	US-10-953-349-9264	Sequence 9264, Ap
12	44	43.6	483	6	US-10-953-349-8401	Sequence 8401, Ap
13	44	43.6	483	6	US-10-953-349-9263	Sequence 9263, Ap
14	44	43.6	485	6	US-10-953-349-8400	Sequence 8400, Ap
15	44	43.6	485	6	US-10-953-349-9262	Sequence 9262, Ap
16	44	43.6	804	7	US-11-293-697-4161	Sequence 4161, Ap
17	44	43.6	1435	6	US-10-196-749-581	Sequence 581, App
18	44	43.6	1743	6	US-10-196-749-451	Sequence 451, App
19	43	42.6	21	7	US-11-144-322-3	Sequence 3, Appli
20	43	42.6	198	6	US-10-449-902-55514	Sequence 55514, A
21	43	42.6	257	6	US-10-953-349-31818	Sequence 31818, A
22	42	41.6	29	1	US-09-949-925-229	Sequence 229, App
23	42	41.6	113	6	US-10-953-349-33908	Sequence 33908, A
24	42	41.6	113	6	US-10-953-349-37356	Sequence 37356, A
25	42	41.6	145	6	US-10-953-349-33907	Sequence 33907, A
26	42	41.6	152	6	US-10-953-349-37355	Sequence 37355, A
27	42	41.6	161	1	US-09-949-925-226	Sequence 226, App
28	42	41.6	217	6	US-10-449-902-39327	Sequence 39327, A
29	42	41.6	414	6	US-10-449-902-32815	Sequence 32815, A
30	42	41.6	414	6	US-10-449-902-37283	Sequence 37283, A
31	42	41.6	414	6	US-10-449-902-46357	Sequence 46357, A
32	41.5	41.1	436	6	US-10-449-902-37829	Sequence 37829, A
33	41	40.6	60	6	US-10-449-902-38433	Sequence 38433, A
34	41	40.6	167	6	US-10-953-349-34493	Sequence 34493, A
35	41	40.6	373	6	US-10-449-902-38114	Sequence 38114, A
36	41	40.6	373	6	US-10-449-902-47991	Sequence 47991, A
37	41	40.6	373	6	US-10-449-902-50488	Sequence 50488, A
38	41	40.6	429	6	US-10-953-349-34644	Sequence 34644, A
39	41	40.6	429	6	US-10-953-349-35589	Sequence 35589, A
40	41	40.6	553	6	US-10-953-349-34643	Sequence 34643, A
41	41	40.6	553	6	US-10-953-349-35588	Sequence 35588, A
√ 42	41	40.6	599	6	US-10-953-349-34642	Sequence 34642, A
43	41	40.6	601	6	US-10-953-349-35587	Sequence 35587, A
44	41	40.6	643	7	US-11-251-673-5	Sequence 5, Appli
45	41	40.6	643	7	US-11-293-697-3832	Sequence 3832, Ap
46	41	40.6	685	7	US-11-293-697-3546	Sequence 3546, Ap
47	41	40.6	720	6	US-10-196-749-170	Sequence 170, App
48	41	40.6	720	7	US-11-101-316-38	Sequence 38, Appl
49	41	40.6	1300	6	US-10-196-749-269	Sequence 269, App
50	40.5	40.1	1066	6	US-10-511-455-2	Sequence 2, Appli

#### ALIGNMENTS

#### RESULT 1

US-10-196-749-525

; Sequence 525, Application US/10196749

; Publication No. US20060094864A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

```

; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C340
; CURRENT APPLICATION NUMBER: US/10/196,749
; CURRENT FILING DATE: 2002-07-16
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 525
; LENGTH: 4440
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-196-749-525

```

```

Query Match          47.5%; Score 48; DB 6; Length 4440;
Best Local Similarity 60.0%; Pred. No. 56;
Matches      9; Conservative      0; Mismatches      6; Indels      0; Gaps      0;

```

```

Qy      3 AAAREACCRECCARA 17
        |||  |||  ||  |
Db      3098 AAACCTACCTTCCGGA 3112

```

```

Search completed: June 19, 2006, 18:01:11
Job time : 20 secs

```

GenCore version 5.1.9  
 Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2006, 17:56:15 ; Search time 125.5 Seconds  
 (without alignments)  
 62.746 Million cell updates/sec

Title: US-10-772-164-1  
 Perfect score: 101  
 Sequence: 1 WEAAAREACCRECCARA 17

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 1000 summaries

Database : Published Applications\_AA\_Main:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match		Length	DB	ID	Description
1	101	100.0		17	3	US-09-973-145-3	Sequence 3, Appli
2	101	100.0		17	3	US-09-880-149-48	Sequence 48, Appl
3	101	100.0		17	3	US-09-880-132-48	Sequence 48, Appl
4	101	100.0		17	3	US-09-813-197-4	Sequence 4, Appli
5	101	100.0		17	4	US-10-126-752-1	Sequence 1, Appli
6	101	100.0		17	4	US-10-174-368A-3	Sequence 3, Appli
7	101	100.0		17	4	US-10-345-281-48	Sequence 48, Appl
8	101	100.0		17	4	US-10-264-127-4	Sequence 4, Appli
9	101	100.0		17	4	US-10-339-712-4	Sequence 4, Appli
10	101	100.0		17	5	US-10-719-523-4	Sequence 4, Appli
11	101	100.0		17	5	US-10-772-164-1	Sequence 1, Appli

12	101	100.0	17	5	US-10-957-433-8	Sequence 8, Appli
13	101	100.0	17	5	US-10-993-568-3	Sequence 3, Appli
14	101	100.0	17	6	US-11-012-853-2	Sequence 2, Appli
15	90	89.1	17	3	US-09-880-149-49	Sequence 49, Appl
16	90	89.1	17	3	US-09-880-132-49	Sequence 49, Appl
17	90	89.1	17	4	US-10-126-752-4	Sequence 4, Appli
18	90	89.1	17	4	US-10-345-281-49	Sequence 49, Appl
19	90	89.1	17	5	US-10-772-164-4	Sequence 4, Appli
20	87	86.1	19	3	US-09-818-875-4368	Sequence 4368, Ap
21	87	86.1	19	4	US-10-260-375A-16	Sequence 16, Appl
22	87	86.1	19	4	US-10-351-662-16	Sequence 16, Appl
23	87	86.1	19	4	US-10-209-787-4368	Sequence 4368, Ap
24	87	86.1	19	4	US-10-307-005-2700	Sequence 2700, Ap
25	87	86.1	19	4	US-10-261-185-4368	Sequence 4368, Ap
26	81	80.2	19	4	US-10-384-918-16	Sequence 16, Appl
27	54	53.5	4277	4	US-10-184-644-439	Sequence 439, App
28	54	53.5	4277	4	US-10-184-634-439	Sequence 439, App
29	53	52.5	189	4	US-10-437-963-149015	Sequence 149015,
30	53	52.5	2974	4	US-10-184-644-521	Sequence 521, App
31	53	52.5	2974	4	US-10-184-634-521	Sequence 521, App
32	50	49.5	28	4	US-10-252-136-14	Sequence 14, Appl
33	50	49.5	28	4	US-10-351-641-231	Sequence 231, App
34	50	49.5	28	4	US-10-267-682-161	Sequence 161, App
35	50	49.5	28	4	US-10-267-748-161	Sequence 161, App
36	50	49.5	152	4	US-10-767-701-60750	Sequence 60750, A
37	50	49.5	284	4	US-10-437-963-199693	Sequence 199693,
38	50	49.5	823	4	US-10-123-155-379	Sequence 379, App
39	50	49.5	823	4	US-10-146-731-379	Sequence 379, App
40	50	49.5	823	4	US-10-140-472-379	Sequence 379, App
41	50	49.5	823	4	US-10-141-761-379	Sequence 379, App
42	50	49.5	823	4	US-10-142-885-379	Sequence 379, App
43	50	49.5	823	4	US-10-158-790-379	Sequence 379, App
44	50	49.5	823	4	US-10-137-871-379	Sequence 379, App
45	50	49.5	823	4	US-10-140-923-379	Sequence 379, App
46	50	49.5	823	4	US-10-141-756-379	Sequence 379, App
47	50	49.5	823	4	US-10-141-759-379	Sequence 379, App
48	50	49.5	823	4	US-10-140-805-379	Sequence 379, App
49	50	49.5	823	4	US-10-140-864-379	Sequence 379, App
50	50	49.5	2012	4	US-10-437-963-204172	Sequence 204172,

#### ALIGNMENTS

#### RESULT 1

US-09-973-145-3

; Sequence 3, Application US/09973145

; Patent No. US20020132248A1

; GENERAL INFORMATION:

; APPLICANT: Rothschild, Kenneth J.

; APPLICANT: Gite, Sadanand

; APPLICANT: Olejnik, Jerzy

; TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins

; FILE REFERENCE: AMBER-06819

; CURRENT APPLICATION NUMBER: US/09/973,145

; CURRENT FILING DATE: 2001-10-09

; PRIOR APPLICATION NUMBER: 09/382,950

; PRIOR FILING DATE: 1999-08-25

; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3  
; LENGTH: 17  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Synthetic  
US-09-973-145-3

Query Match 100.0%; Score 101; DB 3; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1.4e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WEAAAREACCRECCARA 17  
| | | | | | | | | | | | | | |  
Db 1 WEAAAREACCRECCARA 17

Search completed: June 19, 2006, 18:00:47  
Job time : 140.5 secs

GenCore version 5.1.9  
 Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2006, 17:39:41 ; Search time 36.5 Seconds  
 (without alignments)  
 40.768 Million cell updates/sec

Title: US-10-772-164-1  
 Perfect score: 101  
 Sequence: 1 WEAAAREACCRECCARA 17

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 1000 summaries

Database : Issued\_Patents\_AA:\*  
 1: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/5\_COMB.pep:\*  
 2: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/6\_COMB.pep:\*  
 3: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/7\_COMB.pep:\*  
 4: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/H\_COMB.pep:\*  
 5: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
 6: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/RE\_COMB.pep:\*  
 7: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	101	100.0	17	1	US-08-955-206-1	Sequence 1, Appli
2	101	100.0	17	2	US-08-955-050-1	Sequence 1, Appli
3	101	100.0	17	2	US-09-382-950-3	Sequence 3, Appli
4	101	100.0	17	2	US-09-382-736B-4	Sequence 4, Appli
5	101	100.0	17	2	US-09-406-781-48	Sequence 48, Appli
6	101	100.0	17	2	US-09-372-338-1	Sequence 1, Appli
7	101	100.0	17	2	US-09-880-132-48	Sequence 48, Appli
8	101	100.0	17	2	US-10-126-752-1	Sequence 1, Appli
9	101	100.0	17	2	US-09-502-664A-2	Sequence 2, Appli
10	101	100.0	17	2	US-09-813-197-4	Sequence 4, Appli

11	90	89.1	17	1	US-08-955-206-4	Sequence 4, Appli
12	90	89.1	17	2	US-08-955-050-4	Sequence 4, Appli
13	90	89.1	17	2	US-09-406-781-49	Sequence 49, Appl
14	90	89.1	17	2	US-09-372-338-4	Sequence 4, Appli
15	90	89.1	17	2	US-09-880-132-49	Sequence 49, Appl
16	90	89.1	17	2	US-10-126-752-4	Sequence 4, Appli
17	87	86.1	19	2	US-09-818-875-4368	Sequence 4368, Ap
18	56.5	55.9	106	2	US-09-252-991A-24846	Sequence 24846, A
19	54	53.5	245	2	US-09-270-767-35096	Sequence 35096, A
20	54	53.5	245	2	US-09-270-767-50313	Sequence 50313, A
21	53	52.5	365	2	US-09-252-991A-31971	Sequence 31971, A
22	52	51.5	631	2	US-09-252-991A-20063	Sequence 20063, A
23	50	49.5	28	2	US-08-486-099-161	Sequence 161, App
24	50	49.5	28	2	US-08-484-223B-161	Sequence 161, App
25	50	49.5	28	2	US-08-919-597-161	Sequence 161, App
26	50	49.5	28	2	US-08-475-668A-161	Sequence 161, App
27	50	49.5	28	2	US-08-485-551A-161	Sequence 161, App
28	50	49.5	28	2	US-08-471-913A-161	Sequence 161, App
29	50	49.5	28	2	US-08-485-264A-161	Sequence 161, App
30	50	49.5	28	2	US-09-082-279B-231	Sequence 231, App
31	50	49.5	28	2	US-08-474-349A-161	Sequence 161, App
32	50	49.5	28	2	US-09-315-304B-231	Sequence 231, App
33	50	49.5	28	2	US-08-973-952-14	Sequence 14, Appl
34	50	49.5	28	2	US-08-470-896-161	Sequence 161, App
35	50	49.5	28	2	US-08-485-546A-161	Sequence 161, App
36	50	49.5	28	2	US-09-834-784-231	Sequence 231, App
37	50	49.5	28	2	US-09-515-965A-231	Sequence 231, App
38	50	49.5	28	2	US-09-350-641C-231	Sequence 231, App
39	50	49.5	28	2	US-09-350-841A-231	Sequence 231, App
40	50	49.5	28	2	US-08-487-266A-161	Sequence 161, App
41	50	49.5	28	2	US-10-252-136-14	Sequence 14, Appl
42	50	49.5	28	2	US-08-484-741-161	Sequence 161, App
43	50	49.5	62	2	US-09-252-991A-28943	Sequence 28943, A
44	50	49.5	1380	2	US-09-949-016-11688	Sequence 11688, A
45	49.5	49.0	161	2	US-09-252-991A-28201	Sequence 28201, A
46	49	48.5	113	2	US-09-252-991A-19773	Sequence 19773, A
47	48	47.5	162	2	US-09-252-991A-30581	Sequence 30581, A
48	46	45.5	6	2	US-09-818-875-4385	Sequence 4385, Ap
49	46	45.5	101	2	US-09-199-637A-399	Sequence 399, App
50	46	45.5	129	2	US-09-252-991A-22496	Sequence 22496, A

# ALIGNMENTS

## RESULT 1

US-08-955-206-1

; Sequence 1, Application US/08955206

; Patent No. 5932474

; GENERAL INFORMATION:

; APPLICANT: Tsien, Roger Y.

; APPLICANT: Griffin, B. Albert

; TITLE OF INVENTION: TARGET SEQUENCES FOR SYNTHETIC MOLECULES

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 4225 Executive Square, Suite 1400

; CITY: La Jolla

; STATE: CA

```

; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/955,206
; FILING DATE: 21-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07257/060001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; OTHER INFORMATION: the N-terminus is acetylated and
; OTHER INFORMATION: the C-terminus is amidated
US-08-955-206-1

```

```

Query Match          100.0%; Score 101; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 1e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 WEAAAREACCRECCARA 17
        |||||
Db      1 WEAAAREACCRECCARA 17

```

# RESULT 9

US-09-502-664A-2

```

; Sequence 2, Application US/09502664A
; Patent No. 6831160
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: VALE, Ronald
; APPLICANT: THORN, Kurt
; APPLICANT: COOKE, Roger
; APPLICANT: MATUSKA, Marija
; APPLICANT: NABER, Nariman
; TITLE OF INVENTION: METHOD OF AFFINITY PURIFYING PROTEINS USING MODIFIED BIS-
ARSENICAL
; TITLE OF INVENTION: FLUORESC EIN
; FILE REFERENCE: REGEN1500-1
; CURRENT APPLICATION NUMBER: US/09/502,664A
; CURRENT FILING DATE: 2000-02-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2

```



; LENGTH: 17  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: FlAsH-tag peptide  
US-09-502-664A-2

Query Match 100.0%; Score 101; DB 2; Length 17;  
Best Local Similarity 100.0%; Pred. No. 1e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WEAAAREACCRECCARA 17  
| | | | | | | | | | | | | | | |  
Db 1 WEAAAREACCRECCARA 17

Search completed: June 19, 2006, 17:41:11  
Job time : 46.5 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2006, 17:24:19 ; Search time 210.5 Seconds  
(without alignments)  
36.925 Million cell updates/sec

Title: US-10-772-164-1  
Perfect score: 101  
Sequence: 1 WEAAAREACCRECCARA 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database : A\_Geneseq\_8:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*
- 9: geneseqp2005s:\*
- 10: geneseqp2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	101	100.0	17	2	AAAY05336	Aay05336 Target se
2	101	100.0	17	3	AAB20847	Aab20847 Peptide a
3	101	100.0	17	4	AAB35430	Aab35430 Dye-bind
4	101	100.0	17	4	AAM48100	Aam48100 Fluoresce
5	101	100.0	17	8	ADO06947	Ado06947 FLASH-bin
6	101	100.0	17	9	ADZ76895	Adz76895 RNA-tag f
7	90	89.1	17	2	AAAY05337	Aay05337 Target se

8	90	89.1	17	3	AAB20848	Aab20848	Peptide a
9	87	86.1	19	4	AAM51838	Aam51838	Gene corr
10	87	86.1	19	5	AAU81286	Aau81286	Plasmid e
11	87	86.1	19	5	AAU75749	Aau75749	FLAsH pep
12	87	86.1	19	7	ADB78479	Adb78479	FIAsH pep
13	81	80.2	19	7	ABR84531	Abr84531	FLAsH pep
14	76	75.2	595	8	ADQ76865	Adq76865	Adenosine
15	61	60.4	22	3	AAy88739	Aay88739	Core poly
16	61	60.4	22	4	AAB77094	Aab77094	Core poly
17	61	60.4	22	4	ABB00098	Abb00098	Viral DP1
18	61	60.4	22	4	AAU12647	Aau12647	DP178-lik
19	61	60.4	55	5	ADE01583	Ade01583	Hybrid po
20	56.5	55.9	106	7	ABO76100	Abo76100	Pseudomon
21	53	52.5	365	7	ABO83225	Abo83225	Pseudomon
22	52	51.5	631	7	ABO71317	Abo71317	Pseudomon
23	51	50.5	535	8	ADL70535	Adl70535	Human G-p
24	50	49.5	28	3	AAy88872	Aay88872	Core poly
25	50	49.5	28	4	AAB77227	Aab77227	Core poly
26	50	49.5	28	4	ABB00231	Abb00231	Viral DP1
27	50	49.5	28	4	ABB01704	Abb01704	Viral cor
28	50	49.5	28	4	AAU12780	Aau12780	DP178-lik
29	50	49.5	28	6	ABO10317	Abo10317	HIV-1 BRU
30	50	49.5	30	8	ADT71522	Adt71522	Linker mo
31	50	49.5	32	8	ADT71523	Adt71523	Linker mo
32	50	49.5	35	8	ADT71524	Adt71524	Linker mo
33	50	49.5	62	7	ABO80197	Abo80197	Pseudomon
34	50	49.5	906	8	ADP31344	Adp31344	Human sec
35	50	49.5	1134	8	ADP30647	Adp30647	Human sec
36	49.5	49.0	161	7	ABO79455	Abo79455	Pseudomon
37	49	48.5	113	7	ABO71027	Abo71027	Pseudomon
38	49	48.5	120	2	AAW07542	Aaw07542	Clone 99,
39	49	48.5	918	8	ADP31459	Adp31459	Human sec
40	49	48.5	1626	8	ADP31008	Adp31008	Human sec
41	48	47.5	126	2	AAW98909	Aaw98909	Mouse IMC
42	48	47.5	131	2	AAW98908	Aaw98908	Mouse IMC
43	48	47.5	131	7	ADE25527	Ade25527	Mouse SLP
44	48	47.5	131	7	ADF28912	Adf28912	Mouse SLP
45	48	47.5	131	9	ADX02863	Adx02863	Murine an
46	48	47.5	131	10	AEF81210	Aef81210	Spotted s
47	48	47.5	146	8	ADQ59487	Adq59487	Human can
48	48	47.5	146	9	ADZ13856	Adz13856	Murine ca
49	48	47.5	162	7	ABO81835	Abo81835	Pseudomon
50	48	47.5	1305	8	ADP31389	Adp31389	Human sec

# ALIGNMENTS

## RESULT 1

AAy05336

ID AAY05336 standard; peptide; 17 AA.

XX

AC AAY05336;

XX

DT 29-JUN-1999 (first entry)

XX

DE Target sequence peptide, SEQ ID NO. 1.

XX

KW Biarsenical compound; alpha-helix peptide; polypeptide purification;  
 KW immunoassay; crosslinking agent.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9921013-A1.  
 XX  
 PD 29-APR-1999.  
 XX  
 PF 21-OCT-1998; 98WO-US022363.  
 XX  
 PR 21-OCT-1997; 97US-00955050.  
 PR 21-OCT-1997; 97US-00955206.  
 PR 21-OCT-1997; 97US-00955859.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Tsien RY, Griffin AB;  
 XX  
 DR WPI; 1999-288410/24.  
 XX  
 PT Biarsenical compounds that react specifically with cysteine residues.  
 XX  
 PS Claim 10; Page 41; 77pp; English.  
 XX  
 CC This sequence represents a target alpha-helix sequence for the  
 CC biarsenical compounds (BC) of the invention, which are able to react  
 CC specifically with cysteine residues in a target sequence to generate a  
 CC detectable signal. The BCs are used: (i) as labels that allow  
 CC identification of carrier molecules, e.g. in polypeptide purification,  
 CC immunoassays or other chemical or biological assays, including labelling  
 CC in vivo, e.g. to identify, locate or quantify polypeptides or nucleic  
 CC acids); (ii) for attaching a polypeptide to a solid substrate; or (iii)  
 CC to induce a polypeptide domain to adopt a more nearly alpha-helical form,  
 CC e.g. a conformation that can bind a drug. Tetra-arsenical compounds  
 CC derived from the BCs are used to crosslink two binding partners, e.g. to  
 CC study the effect of dimerisation on signal transduction. The BCs react  
 CC specifically with Cys-containing targets, and can be engineered to have  
 CC particular properties, especially ability to cross a biological membrane  
 CC and absence of any self-fluorescence. Both the BC and its target sequence  
 CC are small, and BC binding between them is reversible, e.g. by treatment  
 CC with a dithiol. Particularly, the BC becomes fluorescent when bound to  
 CC its target, but with a significant red-shift from the fluorescence of  
 CC fluorescein, allowing detection with very low background  
 XX  
 SQ Sequence 17 AA;

Query Match 100.0%; Score 101; DB 2; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 1.8e-05;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WEAAAREACCRECCARA 17  
 |||||  
 Db 1 WEAAAREACCRECCARA 17  
 Search completed: June 19, 2006, 17:31:56  
 Job time : 238.5 secs

OM protein - protein search, using sw model

Run on: June 19, 2006, 17:24:38 ; Search time 203 Seconds  
 (without alignments)  
 77.464 Million cell updates/sec

Title: US-10-772-164-4  
 Perfect score: 94  
 Sequence: 1 AEAAAREACCRECCARA 17

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 1000 summaries

Database : UniProt\_7.2:\*  
 1: uniprot\_sprot:\*  
 2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result		%	Query				
No.	Score	Match	Length	DB	ID	Description	
1	51	54.3	278	2	Q6IGH0_DROME	Q6igh0 drosophila	
2	51	54.3	502	2	Q9BGM9_9MAMM	Q9bgm9 tachyglossu	
3	51	54.3	1408	2	Q381X2_9TRYP	Q381x2 trypanosoma	
4	50	53.2	491	2	Q4T2B4_TETNG	Q4t2b4 tetraodon n	
5	50	53.2	589	2	Q2JB13_9ACTO	Q2jb13 frankia sp.	
6	49.5	52.7	602	2	Q75NZ5_CHLRE	Q75nz5 chlamydomon	
7	49	52.1	62	2	Q4PN38_IXOSC	Q4pn38 ixodes scap	
8	49	52.1	157	1	VE6_HPVI2	P36803 human papil	
9	49	52.1	1067	2	Q4QFE4_LEIMA	Q4qfe4 leishmania	
10	48	51.1	358	2	Q3F8C0_9BURK	Q3f8c0 burkholderi	
11	48	51.1	358	2	Q4BRZ7_BURVI	Q4brz7 burkholderi	
12	47	50.0	168	1	VE6_HPVI21	P28832 human papil	
13	47	50.0	193	1	KR415_HUMAN	Q9byq5 homo sapien	
14	47	50.0	199	2	Q3W2K8_9ACTO	Q3w2k8 frankia sp.	
15	47	50.0	210	1	KRA47_HUMAN	Q9byr0 homo sapien	

16	47	50.0	330	2	Q3E2M8_CHLAU	Q3e2m8 chloroflexu
17	47	50.0	399	2	Q3INU5_NATPD	Q3inu5 natronomona
18	47	50.0	438	2	Q341E3_RHOPA	Q341e3 rhodopseudo
19	47	50.0	878	2	Q5RGI5_BRARE	Q5rgi5 brachydanio
20	47	50.0	1959	1	HANG_DROME	Q9vxg1 drosophila
21	46	48.9	80	1	IBB4_LONCA	P16343 lonchocarpu
22	46	48.9	88	2	Q52509_PSESX	Q52509 pseudomonas
23	46	48.9	95	2	Q4CKP2_TRYCR	Q4ckp2 trypanosoma
24	46	48.9	129	1	KRA56_HUMAN	Q618g9 homo sapien
25	46	48.9	161	2	Q8MZ55_DROME	Q8mz55 drosophila
26	46	48.9	186	1	KRA45_HUMAN	Q9byr2 homo sapien
27	46	48.9	191	2	Q28583_SHEEP	Q28583 ovis aries
28	46	48.9	203	2	Q3VGI5_9SPHN	Q3vgi5 sphingopyxi
29	46	48.9	221	2	Q85299_9POXV	Q85299 orf virus.
30	46	48.9	232	2	Q2I0E2_ORYSA	Q2i0e2 oryza sativ
31	46	48.9	412	2	P91666_DROME	P91666 drosophila
32	46	48.9	441	2	Q6N8X8_RHOPA	Q6n8x8 rhodopseudo
33	46	48.9	465	1	HYIN2_BRAJA	P19922 bradyrhizob
34	46	48.9	533	2	Q4S3Z6_TETNG	Q4s3z6 tetraodon n
35	46	48.9	757	2	Q6PFS4_BRARE	Q6pfs4 brachydanio
36	46	48.9	1033	2	Q4T6W6_TETNG	Q4t6w6 tetraodon n
37	46	48.9	1063	2	Q4TBG6_TETNG	Q4tbg6 tetraodon n
38	45	47.9	100	1	YL053_MIMIV	Q5upc9 mimivirus.
39	45	47.9	117	2	Q76YA2_9CAUD	Q76ya2 aeromonas p
40	45	47.9	140	2	Q5TS12_ANOGA	Q5ts12 anopheles g
41	45	47.9	155	2	Q9PXB1_HPV08	Q9pxb1 human papil
42	45	47.9	157	2	O40617_HPVR7	O40617 human papil
43	45	47.9	165	2	Q9D7P3_MOUSE	Q9d7p3 mus musculu
44	45	47.9	204	1	IP22_CAPAN	O49146 capsicum an
45	45	47.9	216	2	Q3TDH8_MOUSE	Q3tdh8 mus musculu
46	45	47.9	233	2	Q7RZM5_NEUCR	Q7rzm5 neurospora
47	45	47.9	250	2	Q3WIB2_9ACTO	Q3wib2 frankia sp.
48	45	47.9	262	2	Q4U5Z5_CAPAN	Q4u5z5 capsicum an
49	45	47.9	262	2	Q4ZIQ3_CAPAN	Q4ziq3 capsicum an
50	45	47.9	262	2	Q4ZIQ4_CAPAN	Q4ziq4 capsicum an

# ALIGNMENTS

## RESULT 1

Q6IGH0\_DROME

ID Q6IGH0\_DROME PRELIMINARY; PRT; 278 AA.

AC Q6IGH0;

DT 05-JUL-2004, integrated into UniProtKB/TrEMBL.

DT 05-JUL-2004, sequence version 1.

DT 07-FEB-2006, entry version 8.

DE HDC06306.

GN ORFNames=HDC06306;

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RX PubMed=14709175; DOI=10.1186/gb-2003-5-1-r3;

RA Hild M., Beckmann B., Haas S.A., Koch B., Solovyev V., Busold C.,

RA Fellenberg K., Boutros M., Vingron M., Sauer F., Hoheisel J.D.,  
 RA Paro R.;  
 RT "An integrated gene annotation and transcriptional profiling approach  
 RT towards the full gene content of the Drosophila genome.";  
 RL Genome Biol. 5:RESEARCH0003.1-RESEARCH0003.17(2003).  
 CC -!- MISCELLANEOUS: The sequence shown here is derived from an  
 CC EMBL/GenBank/DDBJ third party annotation (TPA) entry.  
 CC -----  
 CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>  
 CC Distributed under the Creative Commons Attribution-NoDerivs License  
 CC -----  
 DR EMBL; BK003796; DAA02494.1; -; Genomic\_DNA.  
 DR InterPro; IPR013032; EGF\_like\_reg.  
 DR PROSITE; PS00022; EGF\_1; UNKNOWN\_1.  
 SQ SEQUENCE 278 AA; 32016 MW; 06E7253102FE5BF1 CRC64;

Query Match 54.3%; Score 51; DB 2; Length 278;  
 Best Local Similarity 87.5%; Pred. No. 44;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 9 CCRECCAR 16  
 ||||| |  
 Db 250 CCRECCCR 257

Search completed: June 19, 2006, 17:39:13  
 Job time : 214 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2006, 17:32:24 ; Search time 24 Seconds  
(without alignments)  
68.154 Million cell updates/sec

Title: US-10-772-164-4  
Perfect score: 94  
Sequence: 1 AEAAREACCCECCARA 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database : PIR\_80:\*  
1: pirl:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		%	Query				
No.	Score	Match	Length	DB	ID	Description	
1	49	52.1	157	2	S36538	E6 protein - human	
2	46	48.9	191	2	I46412	keratin KAP5.4 - s	
3	46	48.9	221	2	C34768	ORF2 protein - Orf	
4	46	48.9	273	2	A43862	29K peripheral mem	
5	46	48.9	465	2	S05311	indoleacetamide hy	
6	46	48.9	571	2	S69210	protein kinase cak	
7	45	47.9	204	2	T08072	proteinase inhibit	
8	45	47.9	370	1	S57347	Ca2+/calmodulin-de	
9	45	47.9	374	1	S50193	Ca2+/calmodulin-de	
10	44	46.8	188	2	JC6547	high sulfur protei	
11	44	46.8	217	2	T33353	hypothetical prote	
12	44	46.8	251	2	AH3413	nitrogen fixation	
13	44	46.8	254	2	B84901	hypothetical prote	



14	44	46.8	496	2	F75257	hypothetical prote
15	44	46.8	689	2	T08988	cadmium-transporti
16	44	46.8	711	2	A85352	cadmium-transporti
17	44	46.8	994	2	A48849	Ca2+-transporting
18	44	46.8	1001	1	PWRBFC	Ca2+-transporting
19	43.5	46.3	126	2	I46489	cysteine-rich hair
20	43.5	46.3	229	2	S60454	glucose starvation
21	43	45.7	26	2	C39414	electron transport
22	43	45.7	156	1	W6WL47	E6 protein - human
23	43	45.7	157	1	W6WL5	E6 protein - human
24	43	45.7	157	1	W6WLB5	E6 protein - human
25	43	45.7	169	1	S18946	ultra high-sulfur
26	43	45.7	186	2	A45910	ultra-high-sulfur
27	43	45.7	233	2	S67947	alkyl hydroperoxid
28	43	45.7	399	2	B24698	formate dehydrogen
29	42.5	45.2	101	2	JQ0877	cyc02 protein prec
30	42	44.7	122	2	JC6548	high sulfur protei
31	42	44.7	166	2	S36485	E6 protein - human
32	42	44.7	223	2	B38346	ultra-high-sulfur
33	42	44.7	230	2	A38346	ultra-high-sulfur
34	42	44.7	327	2	C86452	protein F6N18.11 [
35	42	44.7	619	2	C96714	unknown protein T6
36	42	44.7	860	2	A96717	unknown protein, 4
37	42	44.7	997	2	S33754	glutamate receptor
38	42	44.7	2037	2	T16881	hypothetical prote
39	41	43.6	67	2	T37199	hypothetical prote
40	41	43.6	151	2	S60314	hair keratin cyste
41	41	43.6	155	1	W6WL8	E6 protein - human
42	41	43.6	161	2	S36491	E6 protein - human
43	41	43.6	164	2	T24272	hypothetical prote
44	41	43.6	169	2	T06062	hypothetical prote
45	41	43.6	188	2	T15651	hypothetical prote
46	41	43.6	199	2	T48099	hypothetical prote
47	41	43.6	352	2	S11926	cellulose 1,4-beta
48	41	43.6	369	2	F69407	iron-sulfur cluste
49	41	43.6	452	2	G86170	hypothetical prote
50	41	43.6	508	2	T22836	hypothetical prote

#### ALIGNMENTS

##### RESULT 1

S36538

E6 protein - human papillomavirus type 12

C;Species: human papillomavirus type 12

C;Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 09-Jul-2004

C;Accession: S36538

R;Delius, H.; Hofmann, B.

submitted to the EMBL Data Library, August 1993

A;Description: Primer-directed sequencing of human papillomavirus types.

A;Reference number: S36469

A;Accession: S36538

A;Molecule type: DNA

A;Residues: 1-157 <DEL>

A;Cross-references: UNIPROT:P36803; UNIPARC:UPI00001383B8; EMBL:X74466;

NID:g396910; PIDN:CAA52496.1; PID:g396911

C;Superfamily: papillomavirus E6 protein

C;Keywords: DNA binding; early protein; nucleus; zinc finger

Query Match 52.1%; Score 49; DB 2; Length 157;  
Best Local Similarity 87.5%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 8 ACCRECCA 15  
||| |||  
Db 70 ACCRSCCA 77

Search completed: June 19, 2006, 17:39:37  
Job time : 28 secs

OM protein - protein search, using sw model

Run on: June 19, 2006, 17:56:29 ; Search time 13 Seconds  
 (without alignments)  
 29.497 Million cell updates/sec

Title: US-10-772-164-4  
 Perfect score: 94  
 Sequence: 1 AEAAREACCCECCARA 17

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 96747 seqs, 22556637 residues

Total number of hits satisfying chosen parameters: 96747

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 1000 summaries

Database : Published\_Applications\_AA\_New:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*
- 8: /EMC\_Celerra\_SIDS3/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	49	52.1	1129	6	US-10-527-411-42	Sequence 42, Appl
2	49	52.1	1129	6	US-10-527-411-48	Sequence 48, Appl
3	49	52.1	1129	6	US-10-527-411-52	Sequence 52, Appl
4	49	52.1	1129	6	US-10-527-411-56	Sequence 56, Appl
5	49	52.1	1132	6	US-10-527-411-46	Sequence 46, Appl
6	48	51.1	1894	6	US-10-196-749-97	Sequence 97, Appl
7	48	51.1	4440	6	US-10-196-749-525	Sequence 525, App
8	47	50.0	1435	6	US-10-196-749-581	Sequence 581, App
9	46	48.9	1743	6	US-10-196-749-451	Sequence 451, App

10	44	46.8	257	6	US-10-953-349-31818	Sequence 31818, A
11	44	46.8	449	6	US-10-953-349-8402	Sequence 8402, Ap
12	44	46.8	449	6	US-10-953-349-9264	Sequence 9264, Ap
13	44	46.8	483	6	US-10-953-349-8401	Sequence 8401, Ap
14	44	46.8	483	6	US-10-953-349-9263	Sequence 9263, Ap
15	44	46.8	485	6	US-10-953-349-8400	Sequence 8400, Ap
16	44	46.8	485	6	US-10-953-349-9262	Sequence 9262, Ap
17	43	45.7	21	7	US-11-144-322-3	Sequence 3, Appli
18	43	45.7	198	6	US-10-449-902-55514	Sequence 55514, A
19	43	45.7	1300	6	US-10-196-749-269	Sequence 269, App
20	42	44.7	29	1	US-09-949-925-229	Sequence 229, App
21	42	44.7	161	1	US-09-949-925-226	Sequence 226, App
22	42	44.7	217	6	US-10-449-902-39327	Sequence 39327, A
23	42	44.7	1776	6	US-10-933-854-3	Sequence 3, Appli
24	41.5	44.1	113	6	US-10-953-349-33908	Sequence 33908, A
25	41.5	44.1	113	6	US-10-953-349-37356	Sequence 37356, A
26	41.5	44.1	145	6	US-10-953-349-33907	Sequence 33907, A
27	41.5	44.1	152	6	US-10-953-349-37355	Sequence 37355, A
28	41	43.6	167	6	US-10-953-349-34493	Sequence 34493, A
29	41	43.6	429	6	US-10-953-349-34644	Sequence 34644, A
30	41	43.6	429	6	US-10-953-349-35589	Sequence 35589, A
31	41	43.6	553	6	US-10-953-349-34643	Sequence 34643, A
32	41	43.6	553	6	US-10-953-349-35588	Sequence 35588, A
33	41	43.6	599	6	US-10-953-349-34642	Sequence 34642, A
34	41	43.6	601	6	US-10-953-349-35587	Sequence 35587, A
35	41	43.6	643	7	US-11-251-673-5	Sequence 5, Appli
36	41	43.6	643	7	US-11-293-697-3832	Sequence 3832, Ap
37	41	43.6	685	7	US-11-293-697-3546	Sequence 3546, Ap
38	41	43.6	720	6	US-10-196-749-170	Sequence 170, App
39	41	43.6	720	7	US-11-101-316-38	Sequence 38, Appl
40	40.5	43.1	1066	6	US-10-511-455-2	Sequence 2, Appli
41	40	42.6	181	6	US-10-953-349-10362	Sequence 10362, A
42	40	42.6	183	6	US-10-449-902-30401	Sequence 30401, A
43	40	42.6	183	6	US-10-449-902-45055	Sequence 45055, A
44	40	42.6	183	6	US-10-449-902-51021	Sequence 51021, A
45	40	42.6	201	6	US-10-953-349-3609	Sequence 3609, Ap
46	40	42.6	227	6	US-10-449-902-39040	Sequence 39040, A
47	40	42.6	282	7	US-11-293-697-3671	Sequence 3671, Ap
48	40	42.6	306	6	US-10-953-349-3608	Sequence 3608, Ap
49	40	42.6	331	6	US-10-953-349-3607	Sequence 3607, Ap
50	40	42.6	520	6	US-10-449-902-43105	Sequence 43105, A

# ALIGNMENTS

## RESULT 1

US-10-527-411-42

; Sequence 42, Application US/10527411

; Publication No. US20060110410A1

; GENERAL INFORMATION:

; APPLICANT: Shone, Clifford

; APPLICANT: Foster, Keith Alan

; APPLICANT: Chaddock, John

; APPLICANT: Marks, Philip

; APPLICANT: Sutton, J. Mark

; APPLICANT: Stancombe, Patrick

; APPLICANT: Wayne, Jonathan

; TITLE OF INVENTION: Recombinant Toxin Fragments

; FILE REFERENCE: 1581.0130005  
; CURRENT APPLICATION NUMBER: US/10/527,411  
; CURRENT FILING DATE: 2005-03-11  
; PRIOR APPLICATION NUMBER: PCT/GB2003/003824  
; PRIOR FILING DATE: 2003-09-12  
; PRIOR APPLICATION NUMBER: US 10/241,596  
; PRIOR FILING DATE: 2002-09-12  
; NUMBER OF SEQ ID NOS: 175  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 42  
; LENGTH: 1129  
; TYPE: PRT  
; ORGANISM: Clostridium botulinum  
US-10-527-411-42

Query Match 52.1%; Score 49; DB 6; Length 1129;  
Best Local Similarity 58.8%; Pred. No. 14;  
Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AEAAAAREACCRECCARA 17  
|||:| :| :|  
Db 872 AEAAAKEAAAKEAAAKA 888

Search completed: June 19, 2006, 18:01:13  
Job time : 15 secs

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2006, 17:56:15 ; Search time 125.5 Seconds  
(without alignments)  
62.746 Million cell updates/sec

Title: US-10-772-164-4  
Perfect score: 94  
Sequence: 1 AEEAAAREACCCECCARA 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database : Published\_Applications\_AA\_Main:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	94	100.0	17	3	US-09-880-149-49
2	94	100.0	17	3	US-09-880-132-49
3	94	100.0	17	4	US-10-126-752-4
4	94	100.0	17	4	US-10-345-281-49
5	94	100.0	17	5	US-10-772-164-4
6	90	95.7	17	3	US-09-973-145-3
7	90	95.7	17	3	US-09-880-149-48
8	90	95.7	17	3	US-09-880-132-48
9	90	95.7	17	3	US-09-813-197-4
10	90	95.7	17	4	US-10-126-752-1
11	90	95.7	17	4	US-10-174-368A-3

12	90	95.7	17	4	US-10-345-281-48	Sequence 48, Appl
13	90	95.7	17	4	US-10-264-127-4	Sequence 4, Appli
14	90	95.7	17	4	US-10-339-712-4	Sequence 4, Appli
15	90	95.7	17	5	US-10-719-523-4	Sequence 4, Appli
16	90	95.7	17	5	US-10-772-164-1	Sequence 1, Appli
17	90	95.7	17	5	US-10-957-433-8	Sequence 8, Appli
18	90	95.7	17	5	US-10-993-568-3	Sequence 3, Appli
19	90	95.7	17	6	US-11-012-853-2	Sequence 2, Appli
20	87	92.6	19	3	US-09-818-875-4368	Sequence 4368, Ap
21	87	92.6	19	4	US-10-260-375A-16	Sequence 16, Appl
22	87	92.6	19	4	US-10-351-662-16	Sequence 16, Appl
23	87	92.6	19	4	US-10-209-787-4368	Sequence 4368, Ap
24	87	92.6	19	4	US-10-307-005-2700	Sequence 2700, Ap
25	87	92.6	19	4	US-10-261-185-4368	Sequence 4368, Ap
26	81	86.2	19	4	US-10-384-918-16	Sequence 16, Appl
27	54	57.4	4277	4	US-10-184-644-439	Sequence 439, App
28	54	57.4	4277	4	US-10-184-634-439	Sequence 439, App
29	53	56.4	2974	4	US-10-184-644-521	Sequence 521, App
30	53	56.4	2974	4	US-10-184-634-521	Sequence 521, App
31	52	55.3	2076	4	US-10-184-644-409	Sequence 409, App
32	52	55.3	2076	4	US-10-184-634-409	Sequence 409, App
33	51	54.3	2586	4	US-10-063-685-7	Sequence 7, Appli
34	51	54.3	2623	4	US-10-123-155-451	Sequence 451, App
35	51	54.3	2623	4	US-10-146-731-451	Sequence 451, App
36	51	54.3	2623	4	US-10-140-472-451	Sequence 451, App
37	51	54.3	2623	4	US-10-141-761-451	Sequence 451, App
38	51	54.3	2623	4	US-10-142-885-451	Sequence 451, App
39	51	54.3	2623	4	US-10-158-790-451	Sequence 451, App
40	51	54.3	2623	4	US-10-137-871-451	Sequence 451, App
41	51	54.3	2623	4	US-10-140-923-451	Sequence 451, App
42	51	54.3	2623	4	US-10-141-756-451	Sequence 451, App
43	51	54.3	2623	4	US-10-141-759-451	Sequence 451, App
44	51	54.3	2623	4	US-10-140-805-451	Sequence 451, App
45	51	54.3	2623	4	US-10-140-864-451	Sequence 451, App
46	51	54.3	4640	4	US-10-184-644-75	Sequence 75, Appl
47	51	54.3	4640	4	US-10-184-634-75	Sequence 75, Appl
48	50	53.2	28	4	US-10-252-136-14	Sequence 14, Appl
49	50	53.2	28	4	US-10-351-641-231	Sequence 231, App
50	50	53.2	28	4	US-10-267-682-161	Sequence 161, App

#### ALIGNMENTS

#### RESULT 1

US-09-880-149-49

; Sequence 49, Application US/09880149

; Patent No. US20020146843A1

; GENERAL INFORMATION:

; APPLICANT: Kenten, John

; APPLICANT: Roberts, Steven

; TITLE OF INVENTION: CONTROLLING PROTEIN LEVELS IN EUCARYOTIC ORGANISMS

; FILE REFERENCE: 2757-5

; CURRENT APPLICATION NUMBER: US/09/880,149

; CURRENT FILING DATE: 2001-06-14

; PRIOR APPLICATION NUMBER: 09/406,781

; PRIOR FILING DATE: 1999-09-28

; PRIOR APPLICATION NUMBER: 60/119,851

; PRIOR FILING DATE: 1999-02-12  
; NUMBER OF SEQ ID NOS: 67  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 49  
; LENGTH: 17  
; TYPE: PRT  
; ORGANISM: Unknown Organism  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: example peptide  
US-09-880-149-49

Query Match 100.0%; Score 94; DB 3; Length 17;  
Best Local Similarity 100.0%; Pred. No. 9.3e-05;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAAAREACCRECCARA 17  
| | | | | | | | | | | | | | | |  
Db 1 AEAAAREACCRECCARA 17

Search completed: June 19, 2006, 18:00:54  
Job time : 132.5 secs



GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2006, 17:39:41 ; Search time 36.5 Seconds  
(without alignments)  
40.768 Million cell updates/sec

Title: US-10-772-164-4  
Perfect score: 94  
Sequence: 1 AEAAREACCCECCARA 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database : Issued\_Patents\_AA:\*

- 1: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/5\_COMB.pep:\*
- 2: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/6\_COMB.pep:\*
- 3: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/7\_COMB.pep:\*
- 4: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/H\_COMB.pep:\*
- 5: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/PCTUS\_COMB.pep:\*
- 6: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/RE\_COMB.pep:\*
- 7: /EMC\_Celerra\_SIDS3/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	94	100.0	17	1	US-08-955-206-4	Sequence 4, Appli
2	94	100.0	17	2	US-08-955-050-4	Sequence 4, Appli
3	94	100.0	17	2	US-09-406-781-49	Sequence 49, Appl
4	94	100.0	17	2	US-09-372-338-4	Sequence 4, Appli
5	94	100.0	17	2	US-09-880-132-49	Sequence 49, Appl
6	94	100.0	17	2	US-10-126-752-4	Sequence 4, Appli
7	90	95.7	17	1	US-08-955-206-1	Sequence 1, Appli
8	90	95.7	17	2	US-08-955-050-1	Sequence 1, Appli
9	90	95.7	17	2	US-09-382-950-3	Sequence 3, Appli
10	90	95.7	17	2	US-09-382-736B-4	Sequence 4, Appli

11	90	95.7	17	2	US-09-406-781-48	Sequence 48, Appl
12	90	95.7	17	2	US-09-372-338-1	Sequence 1, Appli
13	90	95.7	17	2	US-09-880-132-48	Sequence 48, Appl
14	90	95.7	17	2	US-10-126-752-1	Sequence 1, Appli
15	90	95.7	17	2	US-09-502-664A-2	Sequence 2, Appli
16	90	95.7	17	2	US-09-813-197-4	Sequence 4, Appli
17	87	92.6	19	2	US-09-818-875-4368	Sequence 4368, Ap
18	54	57.4	245	2	US-09-270-767-35096	Sequence 35096, A
19	54	57.4	245	2	US-09-270-767-50313	Sequence 50313, A
20	52.5	55.9	161	2	US-09-252-991A-28201	Sequence 28201, A
21	50	53.2	28	2	US-08-486-099-161	Sequence 161, App
22	50	53.2	28	2	US-08-484-223B-161	Sequence 161, App
23	50	53.2	28	2	US-08-919-597-161	Sequence 161, App
24	50	53.2	28	2	US-08-475-668A-161	Sequence 161, App
25	50	53.2	28	2	US-08-485-551A-161	Sequence 161, App
26	50	53.2	28	2	US-08-471-913A-161	Sequence 161, App
27	50	53.2	28	2	US-08-485-264A-161	Sequence 161, App
28	50	53.2	28	2	US-09-082-279B-231	Sequence 231, App
29	50	53.2	28	2	US-08-474-349A-161	Sequence 161, App
30	50	53.2	28	2	US-09-315-304B-231	Sequence 231, App
31	50	53.2	28	2	US-08-973-952-14	Sequence 14, Appl
32	50	53.2	28	2	US-08-470-896-161	Sequence 161, App
33	50	53.2	28	2	US-08-485-546A-161	Sequence 161, App
34	50	53.2	28	2	US-09-834-784-231	Sequence 231, App
35	50	53.2	28	2	US-09-515-965A-231	Sequence 231, App
36	50	53.2	28	2	US-09-350-641C-231	Sequence 231, App
37	50	53.2	28	2	US-09-350-841A-231	Sequence 231, App
38	50	53.2	28	2	US-08-487-266A-161	Sequence 161, App
39	50	53.2	28	2	US-10-252-136-14	Sequence 14, Appl
40	50	53.2	28	2	US-08-484-741-161	Sequence 161, App
41	49.5	52.7	106	2	US-09-252-991A-24846	Sequence 24846, A
42	49	52.1	113	2	US-09-252-991A-19773	Sequence 19773, A
43	49	52.1	1497	2	US-09-060-854B-2	Sequence 2, Appli
44	49	52.1	1497	2	US-09-529-904-3	Sequence 3, Appli
45	47	50.0	197	2	US-09-252-991A-32518	Sequence 32518, A
46	47	50.0	624	2	US-09-270-767-42659	Sequence 42659, A
47	46	48.9	6	2	US-09-818-875-4385	Sequence 4385, Ap
48	46	48.9	227	2	US-09-252-991A-25546	Sequence 25546, A
49	45	47.9	228	2	US-09-252-991A-30066	Sequence 30066, A
50	45	47.9	370	1	US-08-878-989-19	Sequence 19, Appl

#### ALIGNMENTS

RESULT 1

US-08-955-206-4

; Sequence 4, Application US/08955206

; Patent No. 5932474

; GENERAL INFORMATION:

; APPLICANT: Tsien, Roger Y.

; APPLICANT: Griffin, B. Albert

; TITLE OF INVENTION: TARGET SEQUENCES FOR SYNTHETIC MOLECULES

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 4225 Executive Square, Suite 1400

; CITY: La Jolla

```

; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/955,206
; FILING DATE: 21-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07257/060001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 17 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-955-206-4

```

```

Query Match          100.0%; Score 94; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 6.3e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 AEAAAAREACCRECCARA 17
        ||||||||||||||||
Db      1 AEAAAAREACCRECCARA 17

```

```

Search completed: June 19, 2006, 17:41:14
Job time : 39.5 secs

```

GenCore version 5.1.9  
Copyright (c) 1993 - 2006 Biocceleration Ltd.

OM protein - protein search, using sw model

Run on: June 19, 2006, 17:24:19 ; Search time 210.5 Seconds  
(without alignments)  
36.925 Million cell updates/sec

Title: US-10-772-164-4  
Perfect score: 94  
Sequence: 1 AEAAREACCCECCARA 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database : A\_Geneseq\_8:\*

- 1: geneseqp1980s:\*
- 2: geneseqp1990s:\*
- 3: geneseqp2000s:\*
- 4: geneseqp2001s:\*
- 5: geneseqp2002s:\*
- 6: geneseqp2003as:\*
- 7: geneseqp2003bs:\*
- 8: geneseqp2004s:\*
- 9: geneseqp2005s:\*
- 10: geneseqp2006s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	94	100.0	17	2	AA05337	Aay05337 Target se
2	94	100.0	17	3	AAB20848	Aab20848 Peptide a
3	90	95.7	17	2	AA05336	Aay05336 Target se
4	90	95.7	17	3	AAB20847	Aab20847 Peptide a
5	90	95.7	17	4	AAB35430	Aab35430 Dye-bindi
6	90	95.7	17	4	AAM48100	Aam48100 Fluoresce
7	90	95.7	17	8	ADO06947	Ado06947 FLASH-bin

8	90	95.7	17	9	ADZ76895	Adz76895	RNA-tag f
9	87	92.6	19	4	AAM51838	Aam51838	Gene corr
10	87	92.6	19	5	AAU81286	Aau81286	Plasmid e
11	87	92.6	19	5	AAU75749	Aau75749	FLAsH pep
12	87	92.6	19	7	ADB78479	Adb78479	FIAsH pep
13	81	86.2	19	7	ABR84531	Abr84531	FLAsH pep
14	80	85.1	595	8	ADQ76865	Adq76865	Adenosine
15	54	57.4	22	3	AAy88739	Aay88739	Core poly
16	54	57.4	22	4	AAB77094	Aab77094	Core poly
17	54	57.4	22	4	ABB00098	Abb00098	Viral DP1
18	54	57.4	22	4	AAU12647	Aau12647	DP178-lik
19	54	57.4	55	5	ADE01583	Ade01583	Hybrid po
20	52.5	55.9	161	7	ABO79455	Abo79455	Pseudomon
21	52	55.3	918	8	ADP31459	Adp31459	Human sec
22	52	55.3	1134	8	ADP30647	Adp30647	Human sec
23	51	54.3	2001	8	ADP31644	Adp31644	Human sec
24	50	53.2	28	3	AAy88872	Aay88872	Core poly
25	50	53.2	28	4	AAB77227	Aab77227	Core poly
26	50	53.2	28	4	ABB00231	Abb00231	Viral DP1
27	50	53.2	28	4	ABB01704	Abb01704	Viral cor
28	50	53.2	28	4	AAU12780	Aau12780	DP178-lik
29	50	53.2	28	6	ABO10317	Abo10317	HIV-1 BRU
30	50	53.2	30	8	ADT71522	Adt71522	Linker mo
31	50	53.2	32	8	ADT71523	Adt71523	Linker mo
32	50	53.2	35	8	ADT71524	Adt71524	Linker mo
33	50	53.2	882	8	ADP31688	Adp31688	Human sec
34	50	53.2	906	8	ADP31344	Adp31344	Human sec
35	50	53.2	990	8	ADP31553	Adp31553	Human sec
36	50	53.2	1224	8	ADP31426	Adp31426	Human sec
37	50	53.2	1305	8	ADP31389	Adp31389	Human sec
38	50	53.2	1665	8	ADP31187	Adp31187	Human sec
39	50	53.2	2187	8	ADP30882	Adp30882	Human sec
40	50	53.2	3201	8	ADP31545	Adp31545	Human sec
41	50	53.2	3390	8	ADP31148	Adp31148	Human sec
42	50	53.2	3447	8	ADP31112	Adp31112	Human sec
43	49.5	52.7	106	7	ABO76100	Abo76100	Pseudomon
44	49	52.1	17	10	AEF64458	Aef64458	Protein t
45	49	52.1	19	9	AEA05059	Aea05059	Bradykini
46	49	52.1	21	8	ADN11693	Adn11693	Peptide l
47	49	52.1	113	7	ABO71027	Abo71027	Pseudomon
48	49	52.1	120	2	AAW07542	Aaw07542	Clone 99,
49	49	52.1	1092	8	ADP31358	Adp31358	Human sec
50	49	52.1	1116	8	ADP31692	Adp31692	Human sec

# ALIGNMENTS

## RESULT 1

AAy05337

ID AAY05337 standard; peptide; 17 AA.

XX

AC AAY05337;

XX

DT 29-JUN-1999 (first entry)

XX

DE Target sequence peptide, SEQ ID NO. 4.

XX

KW Biarsenical compound; alpha-helix peptide; polypeptide purification;

KW immunoassay; crosslinking agent.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9921013-A1.  
 XX  
 PD 29-APR-1999.  
 XX  
 PF 21-OCT-1998; 98WO-US022363.  
 XX  
 PR 21-OCT-1997; 97US-00955050.  
 PR 21-OCT-1997; 97US-00955206.  
 PR 21-OCT-1997; 97US-00955859.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Tsien RY, Griffin AB;  
 XX  
 DR WPI; 1999-288410/24.  
 XX  
 PT Biarsenical compounds that react specifically with cysteine residues.  
 XX  
 PS Claim 10; Page 42; 77pp; English.  
 XX  
 CC This sequence represents a target alpha-helix sequence for the  
 CC biarsenical compounds (BC) of the invention, which are able to react  
 CC specifically with cysteine residues in a target sequence to generate a  
 CC detectable signal. The BCs are used: (i) as labels that allow  
 CC identification of carrier molecules, e.g. in polypeptide purification,  
 CC immunoassays or other chemical or biological assays, including labelling  
 CC in vivo, e.g. to identify, locate or quantify polypeptides or nucleic  
 CC acids); (ii) for attaching a polypeptide to a solid substrate; or (iii)  
 CC to induce a polypeptide domain to adopt a more nearly alpha-helical form,  
 CC e.g. a conformation that can bind a drug. Tetra-arsenical compounds  
 CC derived from the BCs are used to crosslink two binding partners, e.g. to  
 CC study the effect of dimerisation on signal transduction. The BCs react  
 CC specifically with Cys-containing targets, and can be engineered to have  
 CC particular properties, especially ability to cross a biological membrane  
 CC and absence of any self-fluorescence. Both the BC and its target sequence  
 CC are small, and BC binding between them is reversible, e.g. by treatment  
 CC with a dithiol. Particularly, the BC becomes fluorescent when bound to  
 CC its target, but with a significant red-shift from the fluorescence of  
 CC fluorescein, allowing detection with very low background  
 XX  
 SQ Sequence 17 AA;

Query Match 100.0%; Score 94; DB 2; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 0.00011;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AEAAAREACCRECCARA 17  
 |||||  
 Db 1 AEAAAREACCRECCARA 17

Search completed: June 19, 2006, 17:32:01  
 Job time : 215.5 secs